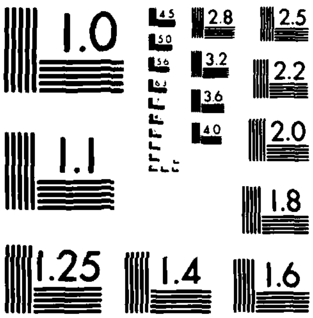
AD-A150 325 A NOTE ABOUT THE STRONG CONVERGENCE OF THE NONPARAMETRIC ESTIMATION OF A..(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS Z FANG SEP 84 1/1

UNCLASSIFIED TR-84-45 AFOSR-TR-85-0005 F49620-82-K-0001 F/G 12/1 NL

END



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A150 325

5

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): TR-84-45	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 85-0005
6a. NAME OF PERFORMING ORGANIZATION: University of Pittsburgh; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): Center for Multivariate Analysis, Pittsburgh PA 15260	7b. ADDRESS (City, State and ZIP Code): Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620-82-K-0001
8c. ADDRESS (City, State and ZIP Code): Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A5	

11. TITLE (Include Security Classification): A NOTE ABOUT THE STRONG CONVERGENCE OF THE NONPARAMETRIC ESTIMATION OF A REGRESSION FUNCTION

12. PERSONAL AUTHOR(S): Zhaoben Fang*

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	SEP 84	7

16. SUPPLEMENTARY NOTATION: *On leave of absence from China University of Science and Technology.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Nonparametric regression; kernel estimation; large sample property.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Consider the regression model ~~$Y_i^* = g(x_i^*) + e_i^*$ i = 1, 2,...n.~~ X_i^*'s that are unordered design variables, g unknown function defined ~~on [0,1]. $\{e_i^*\}$ i.i.d. r.v.~~ with mean 0 and finite moment of order p > 1. The asymptotic behavior of estimator g_n are studied.

cont keywords include

DTIC FILE COPY

DTIC ELECTE FEB 19 1985 S D E

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
MAJ Brian W. Woodruff	(202) 767- 5027	NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

~~AFOSR-TR- 85-0004~~

AFOSR-TR-

A NOTE ABOUT THE STRONG CONVERGENCE
OF THE NONPARAMETRIC ESTIMATION OF
A REGRESSION FUNCTION

Zhaoben Fang

Center for Multivariate Analysis

University of Pittsburgh

85 02 [illegible] 002

A NOTE ABOUT THE STRONG CONVERGENCE OF THE NONPARAMETRIC ESTIMATION OF A REGRESSION FUNCTION

Zhaoben Fang

China University of Science and Technology
and
Center for Multivariate Analysis
University of Pittsburgh

September 1984

Technical Report No. 84-45

Center for Multivariate Analysis
University of Pittsburgh
Room 515 Thackeray Hall
Pittsburgh, PA 15260

A NOTE ABOUT THE STRONG CONVERGENCE OF THE NONPARAMETRIC ESTIMATION OF A REGRESSION FUNCTION

by

Zhaoben Fang

China University of Science and Technology

ABSTRACT

Consider the regression model $Y_i^* = g(x_i^*) + e_i^*$ $i = 1, 2...n$. X_i^*'s are unordered design variables, g unknown function defined on [0,1]. $\{e_i^*\}$ i.i.d r.v with mean 0 and finite moment of order $p > 1$. The asymptotic behavior of estimator g_n are studied.

Key Words: Nonparametric regression, Kernel estimation, large sample property.

We consider the regression model

$$Y_i^* = g(x_i^*) + e_i^* \qquad i = 1, 2, \ldots, n \tag{1}$$

where $x_i^* \in [0,1]$ are unordered design variables. e_i^* are iid random variables with zero mean. $g(x)$ is an unknown function defined on the interval $[0,1]$. The problem is to estimate $g(x)$ through observable variables Y_i^*.

Denote ordered design variables by $x_1 \leq x_2 \leq \ldots \leq x_n$, and define $Y_i = Y_j^*$ and $e_i = e_j^*$ if $x_i = x_j^*$ and $\delta_n = \max_{1 \leq i \leq n+1} (x_i - x_{i-1})$ where $X_0 \equiv 0$, $X_{n+1} \equiv 1$. An estimator proposed by Lin and Cheng [1] is given by

$$g_n(x) = \sum_{i=1}^{n} Y_i \int_{x_{i-1}}^{x_i} a_n^{-1} k\left(\frac{x - z}{a_n}\right) dz \tag{2}$$

where $\{a_n\}$ is a sequence of positive constants converging to 0. as $n \to \infty$, $k(z)$ is a kernel density satisfying:

(i) $k(z) \geq 0$;

(ii) $k(z) = 0$ for $z \notin [-L, L]$ for some positive constant L;

(iii) $\int_{-L}^{L} k(z)dz = 1$.

In fact, it is obtained by the natural estimator

$$h(x) = \begin{cases} Y_1 & x \leq x_1 \\ Y_{i+1} & x_i < x \leq x_{i+1} \quad i = 1, \ldots, n-1 \\ Y_n & x > x_n \end{cases} \tag{3}$$

after smoothing with weight function $a_n^{-1} k[(x - z)/a_n]$. Lin and Cheng [1] discussed the strong convergence of g_n and the rates of uniform convergence under a condition

$E|e^*|^p < \infty$, $p \geq 2$ among other conditions. In this note we relax the restrictions on the moments of e_i^* and improve the results of [1] and [2]. The results are given in the following theorem:

<u>Theorem</u> Assume that k(z) is bounded, $E|e^*|^p < \infty$ for $p > 1$, $\delta_n = 0(n^{-1})$, $a_n = n^{-d}$, for $0 < d < 1 - \frac{1}{p}$. If g(x) is continuous in [0,1], then for all $x \in (0.1)$

$$g_n(x) \to g(x) \qquad \text{a.s.} \tag{4}$$

Proof: Denote $\tilde{e}_i = e_i I(|e_i| \leq i^{1/p})$.

$\tilde{g}_n = \sum_{i=1}^{n} \tilde{e}_i \int_{x_{i-1}}^{x_i} a_n^{-1} k\left(\frac{x - z}{a_n}\right) dz$, where I(·) is an indicator function. Then

$$\begin{aligned} & g_n(x) - g(x) \\ &= g_n(x) - \tilde{g}_n(x) + \tilde{g}_n(x) - E\tilde{g}_n(x) + E\tilde{g}_n(x) - Eg_n(x) + Eg_n(x) - g(x) \\ &\overset{\Delta}{=} I_1 + I_2 + I_3 + I_4. \end{aligned}$$

In the following we prove $I_i \xrightarrow{a.s.} 0$, i = 1, 2 and $I_j \to 0$, j = 3, 4 separately. First, when n is large enough

$$\begin{aligned} I_4 = Eg_n - g &= \sum_{i=1}^{n} g(x_i) \int_{x_{i-1}}^{x_i} a_n^{-1} k\left(\frac{x - z}{a_n}\right) dz - g(x) \\ &= \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} [g(x_i) - g(z)] a_n^{-1} k\left(\frac{x - z}{a_n}\right) dz \\ &\quad + \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} [g(z) - g(x)] a_n^{-1} k\left(\frac{x - z}{a_n}\right) dz \end{aligned} \tag{6}$$

By the uniform continuity of g(x) in [0.1] and the property of k(z), the first term above is less than $\varepsilon \int_0^1 \frac{1}{a_n} k\left(\frac{x - z}{a_n}\right) dz$. where ε is arbitrary positive constant. The second term in (6) is not greater than

$$\sup_{|z-x|<\delta} |g(z)-g(x)| + \frac{1}{\delta} \sup_{|t|>\frac{\delta}{a_n}} |tk(t)| \int_0^1 |g(z)|\,dz$$

$$+ \sup_x |g(x)| \int_{|u|>\frac{\delta}{a_n}} k(u)du.$$

So it is easy to see when $n \to \infty$,

$$I_4 \to 0 \tag{7}$$

by Holder and Chebyshev inequality.

$$|I_3| \le \sum_{i=1}^{n} E\{|e_i| I(|e_i| > i^{1/p})\} \int_{x_{i-1}}^{x_i} a_n^{-1} k\left(\frac{x-z}{a_n}\right) dz$$

$$\le a_n^{-1}\delta_n \|k\| \sum_{i=1}^{n} [P(|e_i| > i^{1/p})]^{1-\frac{1}{p}} (E|e_i|^p)^{\frac{1}{p}}$$

$$\le a_n^{-1}\delta_n \|k\| \sum_{i=1}^{n} \frac{1}{i^{1-\frac{1}{p}}} (E|e_i|^p)^{1-\frac{1}{p}} (E|e_i|^p)^{\frac{1}{p}}$$

$$\le c\, a_n^{-1}\delta_n \sum_{i=1}^{n} \frac{1}{i^{1-\frac{1}{p}}} \le c\, a_n^{-1}\delta_n n^{\frac{1}{p}} \to 0, \qquad n \to \infty \tag{8}$$

where $\|k\| = \sup_x |k(x)|$, c a constant. Notice that if $1 < b \le 2$, we have *$e^z \le 1 + z + |z|^b$ when $z \le 1$, so if r.v $Z_i \le 1$, $EZ_i = 0$ then $E(\exp\{Z_i\}) \le \exp(E|Z_i|^b)$*. Now, take $b < p$, $d_n = \ell_n n \ell_n \ell_n n$, $Z_i = d_n(\tilde{e}_i - E\tilde{e}_i) \int_{x_{i-1}}^{x_i} a_n^{-1} k\left(\frac{x-z}{a_n}\right) dz$. By the choice of d_n and the conditions on δ_n and a_n in the theorem we know that $Z_i \le 1$ when n is large enough. Thus, we have

$$E\left(\prod_{i=1}^{n} \exp\{Z_i\}\right) = \prod_{i=1}^{n} E(\exp\{Z_i\}) \le \prod_{i=1}^{n} \exp(E|Z_i|^b)$$

$$\le \prod_{i=1}^{n} \exp\{d_n (d_n \|k\| \delta_n a_n^{-1})^{b-1} E|e_i|^b \int_{x_{i-1}}^{x_i} a_n^{-1} k\left(\frac{x-z}{a_n}\right) dz\}$$

$$\le \exp\{c\, d_n (d_n \delta_n a_n^{-1})^{b-1}\}$$

For given $\varepsilon > 0$, when n is large enough,

$$P(I_2 > \varepsilon) \leq e^{-d_n \varepsilon} E(\exp\{d_n I_2\})$$

$$= e^{-d_n \varepsilon} E\left(\prod_{i=1}^{n} \exp\{Z_i\}\right) < e^{-\frac{1}{2} d_n \varepsilon}.$$

So, we have

$$\sum_n P(I_2 > \varepsilon) < \sum_n e^{\frac{1}{2} d_n \varepsilon} < \infty$$

by Borel-Cantelli lemma

$$\limsup_{n \to \infty} I_2 \leq 0 \qquad \text{a.s.}$$

Similarly, it is true that

$$\liminf_{n \to \infty} I_2 \geq 0 \qquad \text{a.s.}$$

Therefore

$$I_2 \to 0 \qquad \text{a.s.} \tag{9}$$

Since $E|\, e_i\, |^p < \infty$, we know that $\sum_n P(|\, e_i\, | > j^{\frac{1}{p}}) < \infty$. By Borel-Cantelli lemma we have $P(|\, e_i| > i^{\frac{1}{p}}, \text{ i.o}) = 0$. Hence

$$\sum_{i=1}^{\infty} e_i^2 I(|\, e_i\, | > i^{\frac{1}{p}}) < \infty \quad \text{a.s.} \tag{10}$$

Finally,

$$|\, I_1\, |^2 = |\sum_{i=1}^{n} (e_i - e_i) \int_{x_{i-1}}^{x_i} a_n^{-1} k\left(\frac{x - z}{a_n}\right) dz|^2.$$

By Schwartz inequality,

$$|\, I_1\, |^2 \leq \sum_{i=1}^{n} \left(\int_{x_{i-1}}^{x_i} a_n^{-1}\right) k\left(\frac{x - z}{a_n}\right) dz)^2 \sum_{i=1}^{n} e_i^2 I(|\, e_i\, | > i^{\frac{1}{p}})$$

$$\leq a_n^{-1}\delta_n \, \| k \| \int_0^1 a_n^{-1} k\left(\frac{x-z}{a_n}\right) dz \sum_{i=1}^{\infty} e_i^2 I(| e_i | > i^{\frac{1}{p}})$$

$$\leq c\, a_n^{-1}\delta_n \sum_i e_i^2 I(| e_i | > i^{\frac{1}{p}}).$$

By (10) we have

$$| I_1|^2 \to 0 \quad \text{a.s.} \tag{11}$$

From (5), (7), (8), (9), and (11), the theorem follows.

REFERENCES

[1] CHENG, Kuang-Fu and LIN, Pi-Erh, Z. Wahrscheinlichkeitstheorie venv. Gebiete, 57, 223-233 (1981).

[2] BENEDETTI, J. Roy. Statist. Soc. B39, 243-253, (1977).

END

FILMED

3-85

DTIC